

2010-05

# Research by the Numbers: Assessing the Performance of Three Products for Citation Analysis

Conte, Marisa; Song, Jean; Townsend, Whitney

<https://hdl.handle.net/2027.42/94430>

<http://creativecommons.org/licenses/by/3.0/>



## Objective

Citation analysis projects constitute a significant portion of the research requests submitted to the Taubman Health Sciences Library (THL). Researchers and administrators are increasingly interested in citation analysis as a way to describe research performance or impact, with particular interest on the cited-by and h-index metrics. Several commercial products provide access to citation analysis metrics, but users are often unaware of these features and generally do not understand what the numbers really mean or the context in which they can be appropriately applied. The purpose of this study is to compare the citation analysis functions of Collexis Research Profiles, ISI Web of Science, and Scopus in terms of accuracy, coverage, and overall functionality.

## Background

### Common Citation Metrics

Metric	Definition	Advantage	Disadvantage
<b>Number of Papers</b>	The total number of papers published by a researcher	Measures productivity	Does not measure importance or impact
<b>Number of Citations</b>	The total number of citations that cite a researcher's papers	Measures total impact	May be inflated by big hits
<b>h-index</b>	Distribution of citations to a researcher's publications	Addresses quality and quantity	Misinterpretation: low h-index does not necessarily reflect under-achievement

### Coverage and Data Sources

Resource	h-index	Total Citation counts	Cited By counts
<b>Collexis Research Profiles</b>	ISI Web of Science MEDLINE subset	Total PubMed/MEDLINE content	ISI Web of Science MEDLINE subset
<b>ISI Web of Science</b>	Depth of subscription coverage	Total Web of Science content	Total Web of Science content
<b>Scopus</b>	Scopus database content 1996 - present	Total Scopus database content (100% MEDLINE coverage)	Total Scopus database content (100% MEDLINE coverage)

## Methodology

A random sample of 50 author names was generated from a prior citation analysis request submitted from an administrative unit at the University of Michigan (UM) Health System and conducted by a THL librarian. The 50 author names were run against each of the three designated databases: Collexis Research Profiles, ISI Web of Science, and Scopus using the specific search strategies detailed below. Citation analysis metrics for total number of publications, h-index and total cited-by citations were recorded. For the graphs displayed below, a further random sample of 10 names was selected.

### Search Strategies:

Collexis Research Profiles:

- Using 'By Last Name' search feature, enter author's last name
- Select appropriate author from list
- Record total number of publications and h-index from Researcher Profile
- Manually calculate cited-by citation counts from Researcher Profile

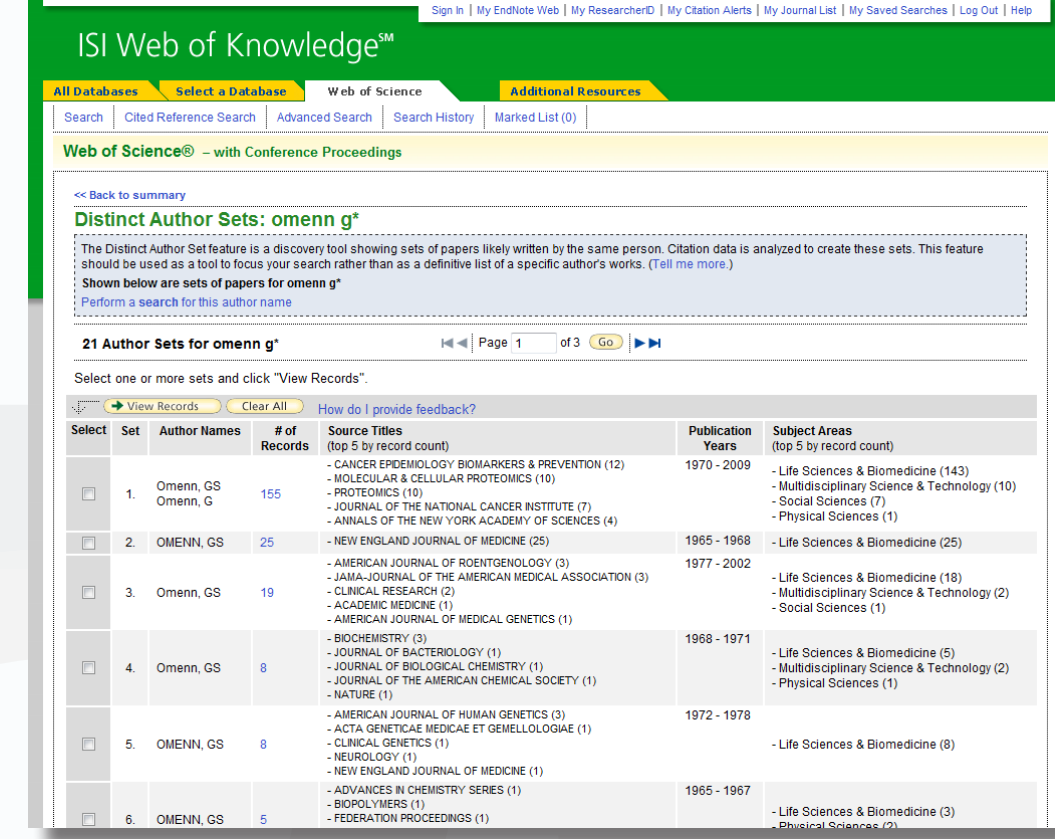
### Collexis Research Profiles



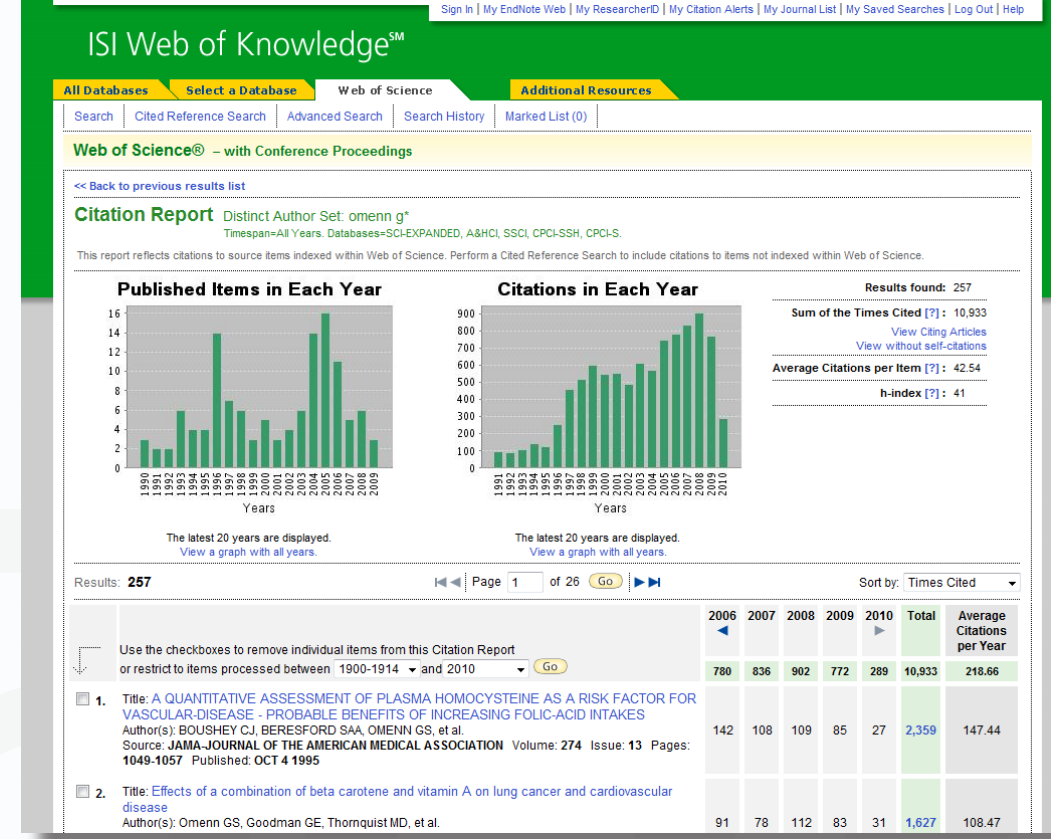
### ISI Web of Science:

- Using 'Author' search field, enter author's last name and first initial with an asterisk
- If further disambiguation is necessary, enter first and middle initial with an asterisk
- Select relevant author sets using the 'Distinct Author Sets' tool
- Use the 'Create Citation Report' tool to generate citation report
- Record total number of publications, h-index, cited-by citation counts

### ISI Web of Science Distinct Author Sets



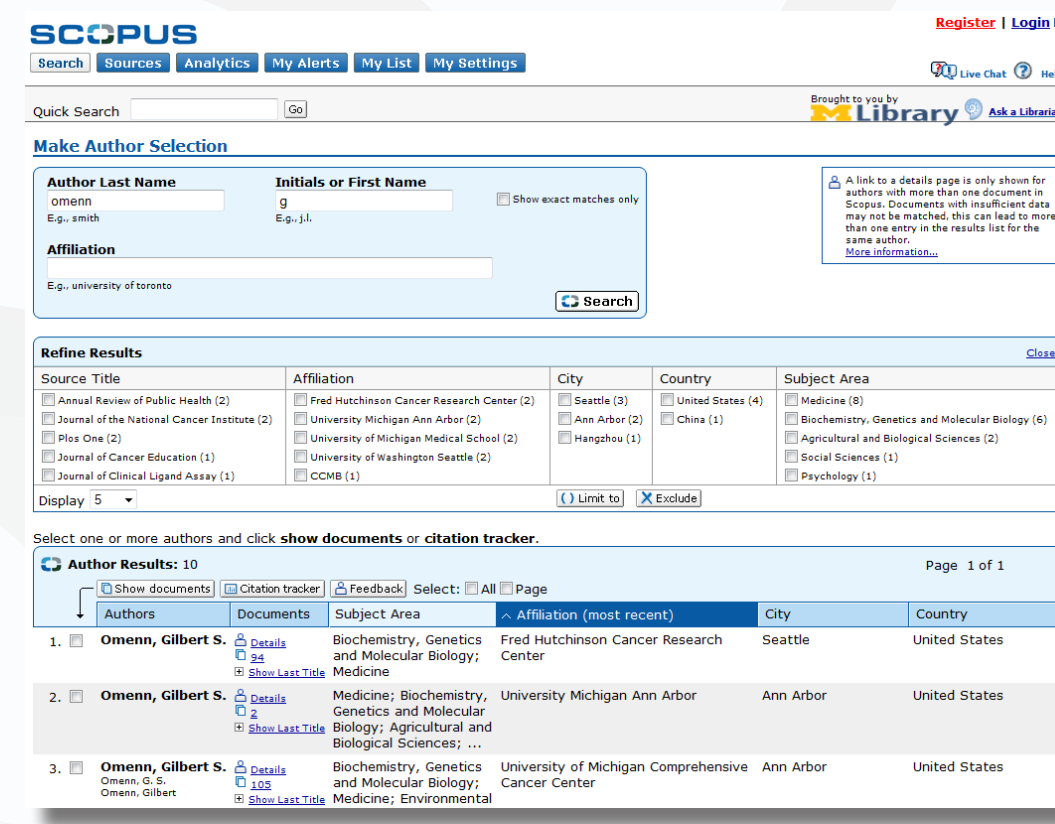
### ISI Web of Science Citation Report



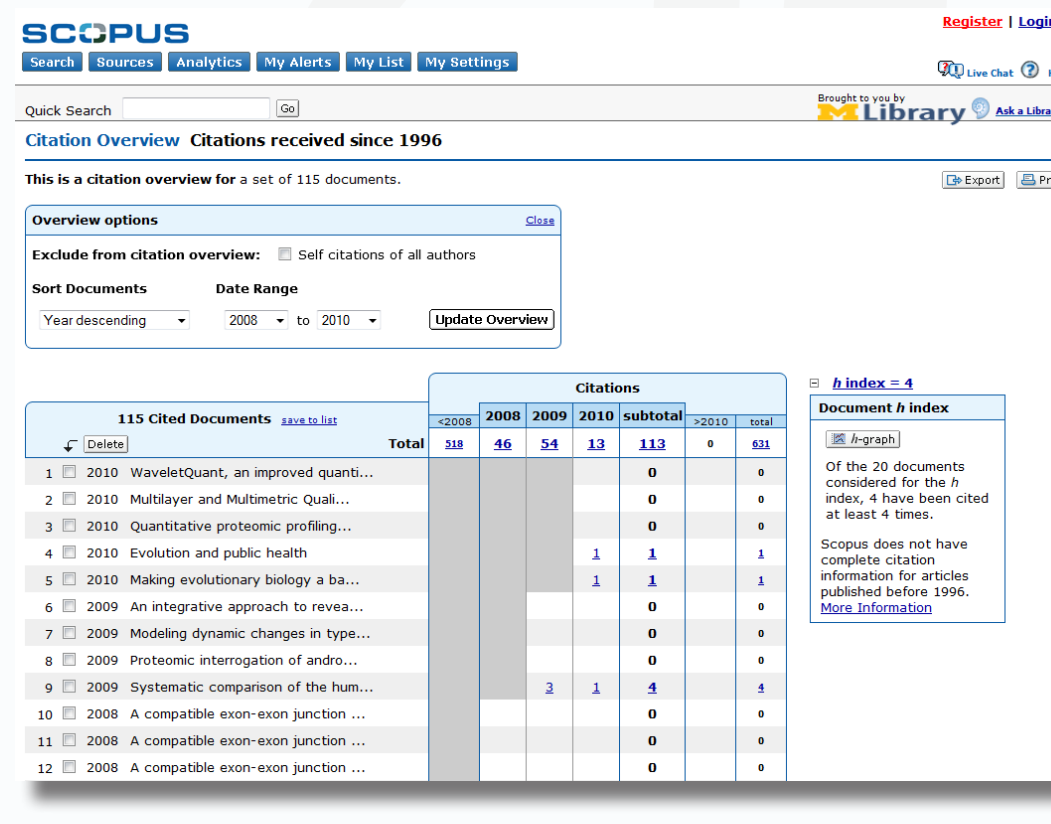
### Scopus:

- Using 'Author Search' feature, enter author's last name and first initial
- If further disambiguation is necessary, enter full first name
- Select relevant records and select 'Citation Tracker' feature
- Record total number of publications, h-index, cited-by citation counts

### Scopus Author Results



### Scopus Citation Overview

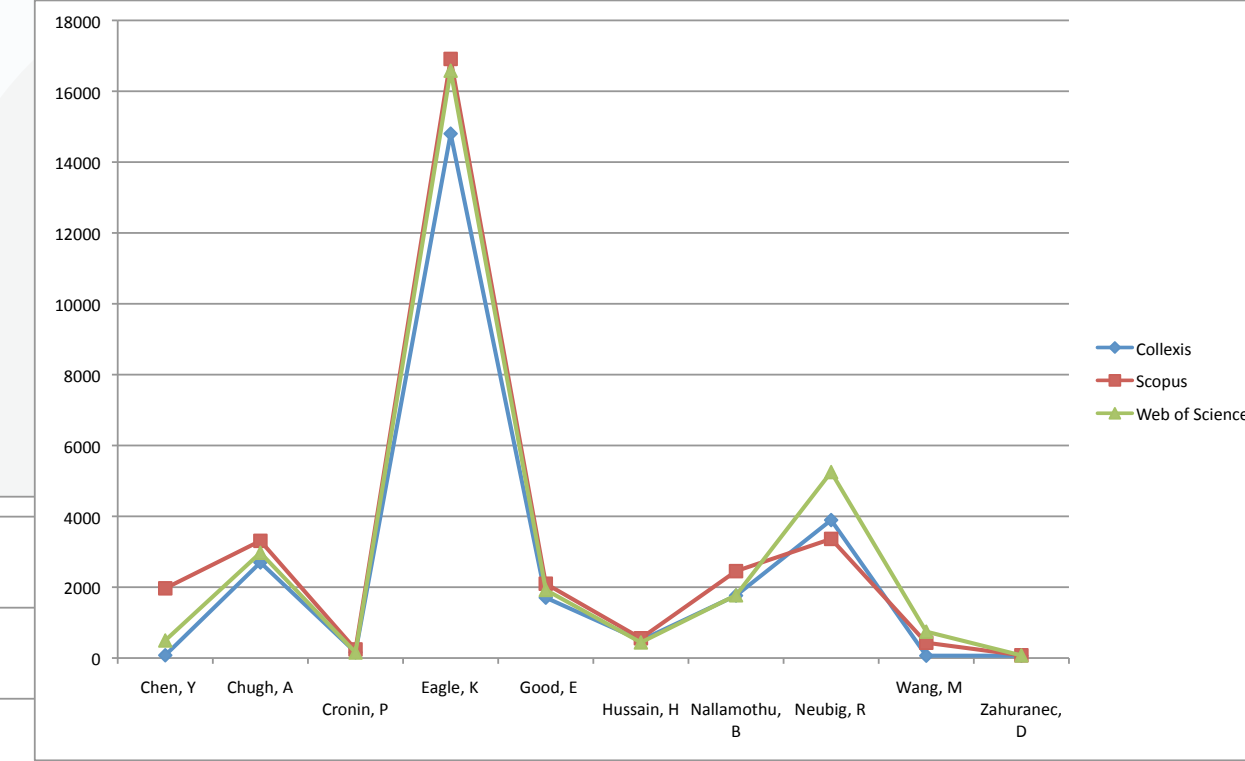


### Selected Results and Discussion Points

- In general, for the sample subset of 10 random researchers used to display the data for this poster:
  - the cited-by metric was higher in Scopus than the other two tools
  - the total publications metric was higher in ISI Web of Science than the other two tools
  - variability in one citation metric was an indicator of variability in the other citation metrics
- Given that ISI Web of Science is the data source for Collexis' citation metrics, the variability between the h-index and cited by citation metrics was greater than expected. It is presumed that this is due to the limitations of the ISI Web of Science author sets, which will be further investigated.

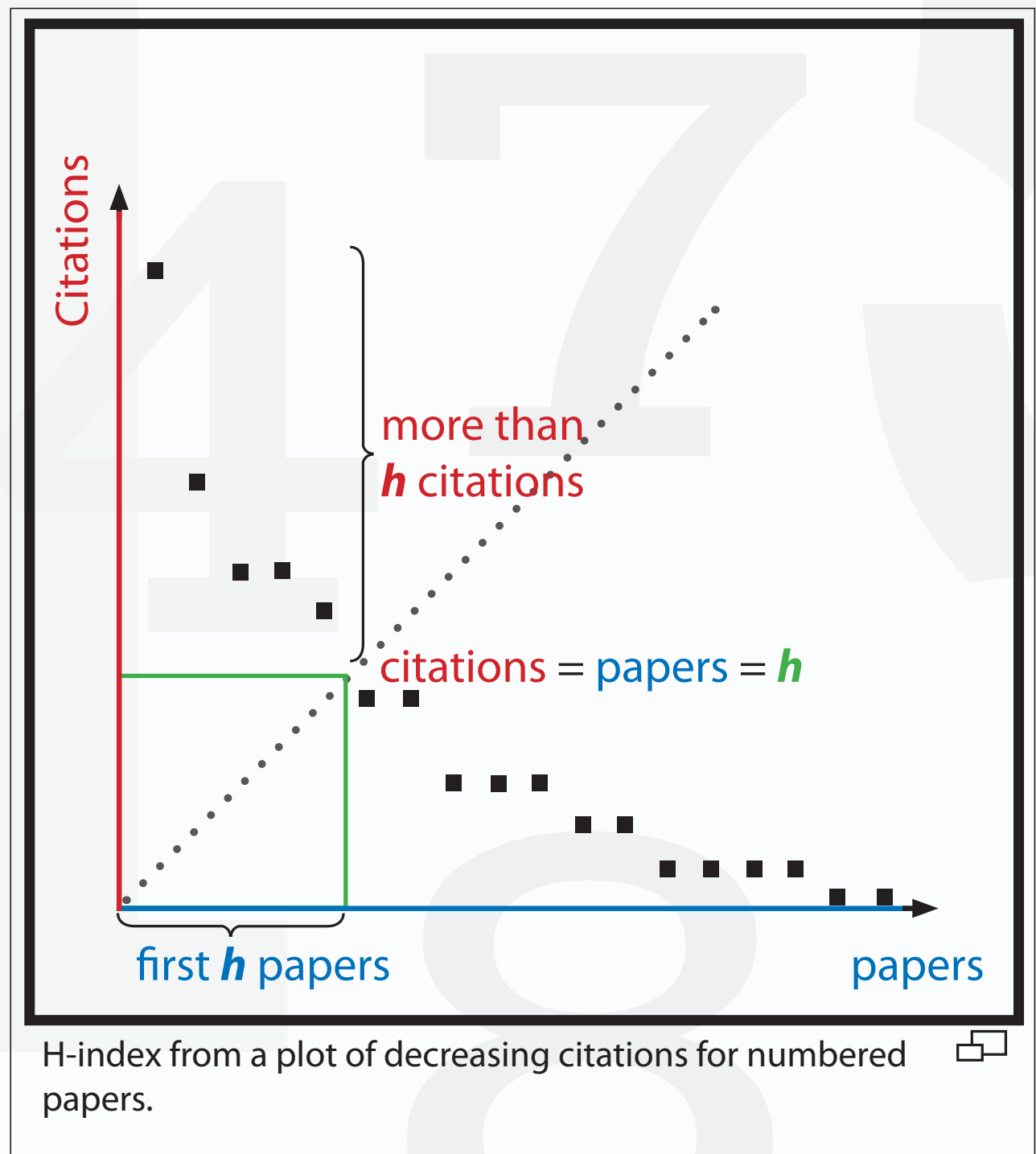
Name	h-index	Total Publications	Cited By
Chen, Y.	4	26	10
Chen, A.	23	26	25
Crowin, P.	6	6	5
Cagle, K.	60	60	63
Gross, E.	20	22	15
Hussain, H.	13	13	11
Nalamothu, B.	23	26	21
Wang, M.	4	7	13
Zahurancu, D.	5	5	5

### Cited by Comparison for Selected Researchers



Total Publications Comparison for Selected Researchers

h-index Comparison for Selected Researchers



H-index from a plot of decreasing citations for numbered papers.

Diagrammatic View of h-index<sup>1</sup>

- x-axis: number of publications
- y-axis: number of citations
- h: point at which number of citations = number of papers

<sup>1</sup><http://en.wikipedia.org/wiki/H-index>

### Citation Analysis Functions Summary

Citation Analysis	Collexis Research Profiles	Scopus	ISI Web of Science
Author disambiguation	UM's instantiation involves time-consuming manual curation	Limited author record marking	Distinct author sets are incomplete; Ambiguity still exists within author sets
Metrics location	Expert Overview and Publications page	Citation Overview (citation tracker button)	Citation Report (create citation report link)
h-index	Limited to ISI Web of Science MEDLINE subset	Different values for the same author depending on database function	Limited to institutional database subscription coverage
Total publications	Time intensive manual calculation	Limited to Scopus database coverage	Limited to ISI Web of Science database coverage
Cited by	Counts must be manually totaled, introducing possibility of error	Limited to Scopus database coverage	Limited to ISI Web of Science database coverage

## Limitations

Search strategies were not intended to be comprehensive or capture a definitive list of each author's research output. Rather, the intent was to utilize search features within each resource for author disambiguation and compare the utility and accuracy of the results.

The different data sources represented in this study each have their own limitations. While none of these limitations represent fatal flaws in the respective products, they are valid concerns or considerations that any end user needs to consider when evaluating the product and its results.

### Collexis Research Profiles:

- UM's current instantiation includes only those authors with a primary appointment to clinical or research tracks at the University of Michigan Medical School (UMMS). While all of the sampled authors met this criteria, the resource is currently not comprehensive for all UM or UMMS faculty.
- UM's current instantiation is manually curated to resolve author name disambiguation issues and validated against faculty curriculum vitae to increase publication count accuracy. However, the h-index and cited-by counts are currently based on ISI Web of Science's MEDLINE subset, which does not represent full MEDLINE.
- There is no way to manually refine a name search within the Research Profiles.
- Cited-by counts are not totaled and need to be manually calculated.

### ISI Web of Science:

- Citation lists using the 'Distinct Author Sets' tool are markedly smaller than citation lists using author last name and first initial. Many citations do not appear to have been grouped into the provided author sets.
- Some common name/initial sets are impossible to disambiguate using the 'Distinct Author Sets' tool. Full author first names are not searchable, making the disambiguation process for common names even more difficult.

### Scopus:

- For name disambiguation issues, a user can select only up to 15 names. This resulted in the exclusion of many individual publications not linked to an author name record.
- Unclear algorithm for calculating h-index using different database features: in several instances, a different h-index was obtained from the same set of articles due to an unexplained difference in the number of articles used to calculate h-index. A query to Scopus Helpdesk was still open at the time the poster was printed.
- During the period of data collection, the Scopus database presented the following error message: "Scopus is currently experiencing issues with the retrieval and display of some search results. We aim to correct this processing error as soon as possible. Please forgive the inconvenience." Communication with the Scopus Helpdesk indicated that the problem was fixed but the message persisted throughout the data collection period.

## Conclusion

The citation analysis functions for each of the three tools have unique strengths and weaknesses. Clearly, author name disambiguation continues to pose a problem, particularly in Scopus and ISI Web of Science. UM's instantiation of the Collexis Research Profiles goes a long way to addressing this issue through manual curation, but this is a time- and resource-intensive solution that may not be appropriate for all institutions. Collexis Research Profiles, however, uses only a subset of citations from ISI Web of Science to calculate its h-index and cited by metrics. While all three products provide readily available citation metric data through their reporting features, end users need to be aware of the limitations of coverage and accuracy when using this information.

## Acknowledgements

\*Funding provided by the Michigan Institute for Clinical Health Research, NIH Grant UL1RR024986.

\*\*Funding provided by the National Center for Integrative Biomedical Informatics, NIH Grant U54DA021519.